Computer-aided approaches reveal trihydroxychroman and pyrazolone derivatives as potential inhibitors of SARS-CoV-2 virus main protease

NOOR ATATREH[1,*]
SHAIMA HASAN[1]
BASSAM R. ALI[2]
MOHAMMAD A. GHATTAS[1,*]

[1] *College of Pharmacy, Al Ain University Abu Dhabi, UAE, P.O. Box 112612*

[2] *Department of Pathology, College of Medicine and Health Sciences United Arab Emirates University Al Ain, UAE*

COVID-19 was declared a pandemic by the World Health Organization (WHO) in March 2020. The disease is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The aim of this study is to target the SARS-CoV-2 virus main protease (M^pro) *via* structure-based virtual screening. Consequently, > 580,000 ligands were processed *via* several filtration and docking steps, then the top 21 compounds were analysed extensively *via* MM-GBSA scoring and molecular dynamic simulations. Interestingly, the top compounds showed favorable binding energies and binding patterns to the protease enzyme, forming interactions with several key residues. Trihydroxychroman and pyrazolone derivatives, SN02 and SN18 ligands, exhibited very promising binding modes along with the best MM-GBSA scoring of -40.9 and -41.2 kcal mol^{-1}, resp. MD simulations of 300 ns for the ligand-protein complexes of SN02 and SN18 affirmed the previously attained results of the potential inhibition activity of these two ligands. These potential inhibitors can be the starting point for further studies to pave way for the discovery of new antiviral drugs for SARS-CoV-2.

Keywords: COVID-19, M^pro, SARS-COV-2, antiviral, virtual screening, docking

The coronavirus disease (COVID-19) is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). This virus has 82 and ~50 % structural similarities with SARS-CoV and MERS-CoV, resp. (1). SARS-CoV-2 genome is a positive sense, single-stranded RNA and the main structure consists of 14 open reading frames (ORFs), which encode 15 non-structural proteins, 4 structural proteins and 8 accessory proteins (2). The most important non-structural protein (nsp) is the main protease (M^pro) which plays an important role in the virus life cycle (3), and its crystal structure has been recently revealed through X-ray crystallography (4).

Although there are no currently FDA-approved medicines to treat COVID-19 disease, there are several potential drugs to inhibit SARS-CoV-2 activity (5). These compounds can

* Correspondence; e-mail: mohammad.ghattas@aau.ac.ae; noor.atatreh@aau.ac.ae

be classified according to their mode of action into inhibitors of the TMPRSS2 serine protease and subsequently preventing the virus from entering the host cell (6), such as camostat mesylate, drugs that block the interaction between the virus and the host angiotensin--converting enzyme II (ACE2), like hydroxychloroquine, or those that inhibit the viral RNA-dependent RNA polymerase (RdRp) protein, such as remdesivir and lopinavir, and finally drugs that inhibit the activity of SARS-CoV-2 M^{pro} protease (4).

Despite these advantages, no efficient treatment is currently available for COVID-19 and the death toll is still rising daily. Thus, the goal of this project is to discover new potential inhibitors of the SARS-CoV-2 M^{pro} enzyme *via* computer-aided drug design.

<div align="center">EXPERIMENTAL</div>

Virtual screening

The M^{pro} crystal structure [PDB code: 6LU7 (4)] was first checked and corrected *via* the MOE software (7). Then, it was processed through Protein Preparation Wizard in the Maestro software (8), where hydrogen atoms and partial charges were added accordingly. Ligand database was obtained from a commercial source (TimTec LLC), then prepared by LigPrep in Maestro and afterwards filtered based on the drug-like rules (rotatable bond ≤ 10, molecular mass ≤ 500, log*P* ≤ 5.0, PSA ≤ 140, HB donor ≤ 5, HB acceptor ≤ 10). The filtered ligand dataset was docked into the M^{pro} active site. The Glide (9) virtual screening workflow (VSW) is a three-step protocol that includes three levels of improved docking precisions: high-throughput virtual screening (HTVS), standard precision (SP), and extra-precision (XP). The top 30 % scoring ligands, HTVS and SP, were selected for the next level of docking. That ended up with 78,783 ligand enumerations that were docked and ranked by GLIDE-XP. Visual inspection was conducted afterwards to check the hits fitting into the target pocket. Finally, 21 ligands belonging to various structural scaffolds and showing convenient binding modes in the M^{pro} active site were selected.

MD simulation and MM-GBSA scoring

MD simulation studies were conducted to further investigate ligand's binding affinities and to inspect ligand's fitting into the target pocket. Two stages of MD simulation were carried out for the selected 21 compounds using the AMBER 18 Force field (10). Ligands were assigned partial charges *via* the Antechamber program (11) using the Generalized Amber Force Field (GAFF) and AM1-BCC (12). The protein structure was assigned partial charges and other parameters using the ff19SB force field. The ligand-protein complex system was built using the xleap module of AmberTools, where it was neutralized by the addition of Na$^+$ counter ions and solvated by a truncated octahedral box of TIP3P water.

The system was then energy minimized using the pmemd program in the AMBER 18 package (10) through two steps: firstly, all solute atoms were restrained with a force constant of 500 kcal mol^{-1} Å^{-2} during minimization, and then the whole system was subjected to minimization without applying any restraints. Using MD simulation, the system was then gradually heated through the NVT ensemble from 0 to 300 K with a 10 kcal mol^{-1} Å^{-2} restraint on ligand atoms over 20 ps. The SHAKE algorithm was employed for all bonds involving hydrogen atoms. Finally, a production MD simulation of 20 ns for all 21 com-

pounds was performed under NPT conditions with a system temperature and pressure fixed to 300 K and of 1.01×10^5 Pa, respectively. MD simulation was continued till 300 ns for compounds with MM-GBSA score ≤ -40 mol^{-1} (SN02 and SN18).

Clustering and PLIF analysis

Clustering was performed using DBSCAN *via* the cpptraj module of AmberTools (14). For each ligand-protein complex, resultant frames from the desired period of the simulation were utilized in the clustering process while skipping every 10th frame. Each system was stripped of ions and solvent molecules. The distance cutoff between points for forming a cluster was set to 3.0 (default value). Protein-ligand interaction fingerprint (PLIF) was conducted for the top clustered configuration of each complex obtained from the MD simulation. The PLIF analysis was generated using the MOE software where the minimum score thresholds for hydrogen bond and arene attraction were set to 1.5 and 1 kcal mol^{-1} (default settings), resp. (7). The created PLIF graph displays the interaction occupancy of all residues in the M^{pro} active site, defined as the percentage of ligands interacting with the given amino acid.

RESULTS AND DISCUSSION

Fig. 1 schematically illustrates in the virtual screening protocol conducted against the SARS-CoV-2 protease enzyme. A commercial ligand library (TimTec company) was first to filter for druglikeness before docking into the M^{pro} active site, reducing the library size from 582,474 ligands to 440,617 ligands. The drug-like library was initially screened using the GLIDE-HTVS docking algorithm. The top 30 % of ligands were re-docked in a slower but more accurate docking algorithm using GLIDE-SP. Finally, docking results were further refined by docking the resultant top 30 % using the extra-precision algorithm (GLIDE-XP) in the Schrödinger software package.

The resultant docked poses of the top 500 ligands were visually inspected inside the M^{pro} binding site based on their fitting into the cavity and their interactions with the surrounding residues. A shortlist of 21 compounds were selected from various chemical clusters and showed to have favourable docking scores ranging from –7.4 to –8.8 kcal mol^{-1} (Table I).

As docking suffers from poor scoring function, more accurate methods such as MM-GBSA calculations are used to re-score docked ligands and predict their binding energies. MM-GBSA scoring method has shown a better correlation (than docking-associated scoring functions) between predicted scores and the actual inhibition constant (K_i) or IC_{50} (13). Consequently, these shortlisted 21 compounds were simulated with the target enzyme for 20 ns, then they were rescored *via* the MM-GBSA method. Interestingly, all 21 compounds have shown low and favourable binding energies ranging from –20 to –41.2 kcal mol^{-1}. Table II lists the MM-GBSA scores along with each component scoring obtained for the best four compounds, having SN02 and SN18 scoring the lowest binding energies of all tested ligands (< –40 kcal mol^{-1}).

It was interesting to check which active residues in the M^{pro} pocket are mostly involved when interacting with the bound ligands. To do that, the most populated binding modes of these 21 compounds were identified through clustering the resulting frames of the 10 ns
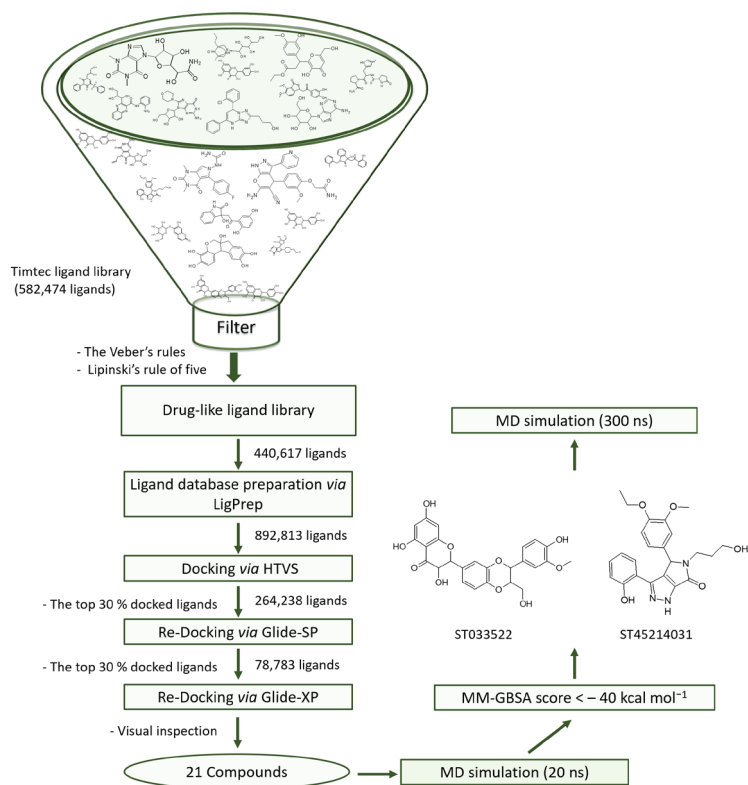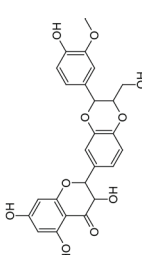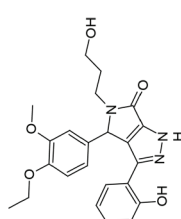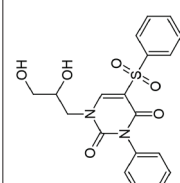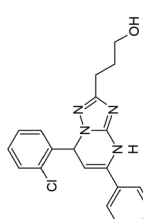
Fig. 1. Virtual screening workflow conducted against the M^pro active site.

Table I. Docking scores of the best compounds selected based on visual inspection

Name	Compound ID	GLIDE-XP score (kcal mol⁻¹)	Name	Compound ID	GLIDE-XP score (kcal mol⁻¹)
SN01	HTS34630	−7.60	SN12	ST403229	−7.97
SN02	ST033522	−8.81	SN13	ST4073948	−7.97
SN03	ST056309	−7.83	SN14	ST4083738	−7.71
SN04	ST057134	−8.09	SN15	ST4092083	−8.53
SN05	ST057145	−7.79	SN16	ST4138931	−7.36
SN06	ST060285	−8.00	SN17	ST4151611	−8.27
SN07	ST069336	−8.18	SN18	ST45214031	−7.63
SN08	ST070937	−7.96	SN19	ST45262616	−7.82
SN09	ST075200	−7.72	SN20	ST50072883	−8.10
SN10	ST075699	−7.82	SN21	ST50892216	−7.85
SN11	ST082430	−7.73			

Table II. MD Simulation results for the best simulated ligands with MM-GBSA score < −35 kcal mol⁻¹

Name	Compd. ID	Structure	MM-GBSA ± SD (kcal mol⁻¹)	VDWAALS ± SD (kcal mol⁻¹)	EGB ± SD (kcal mol⁻¹)	ESURF ± SD (kcal mol⁻¹)	EEL ± SD (kcal mol⁻¹)
SN02	ST033522		−40.9 ± 3.1	−48.8 ± 3.3	55.2 ± 4.1	−6.4 ± 0.3	−40.9 ± 5.8
SN18	ST45214031		−41.2 ± 4.3	−48.2 ± 3.5	50.9 ± 6.7	−5.8 ± 0.3	−38.2 ± 9.9
SN19	ST45262616		−35.8 ± 4.1	−45.8 ± 3.8	56.2 ± 4.8	−5.6 ± 0.4	−40.7 ± 7.3
SN20	ST50072883		−36.3 ± 3.9	−43.9 ± 3.8	30.9 ± 3.1	−5.5 ± 0.3	−17.8 ± 4.9

EEL – electrostatic energy, EGB – polar contribution to the solvation free energy, ESURF – nonpolar contribution to the solvation free area, MM-GBSA – molecular mechanics-generalize born surface area, VDWAALS – van der Waals contribution
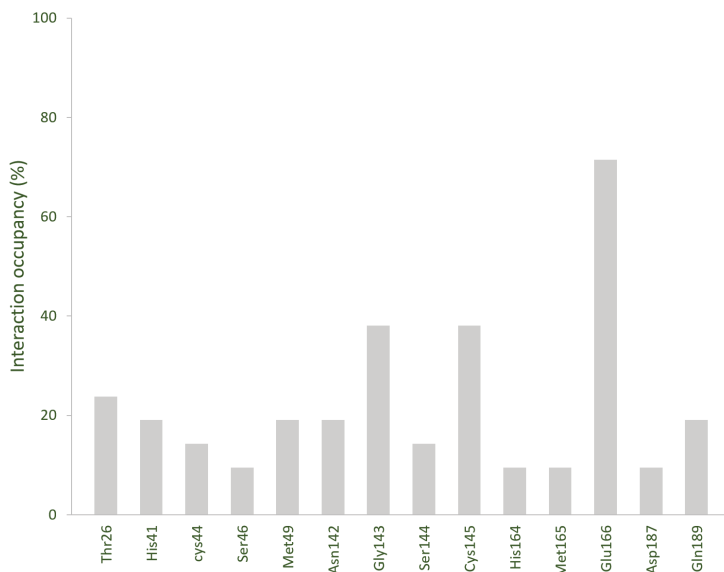
Fig. 2. Protein-ligand interaction fingerprint of the visually selected 21 compounds.

of the MD simulation. As shown in Fig. 2, the most engaged residue in ligand interaction is Glu166 (15 out of 21 ligands), which has been previously identified as one of the important residues for M^{pro} inhibitors binding (4, 15). Other residues to interact with more than 20 % of ligands are Thr 26, Gly143 and Cys145; the latter is the nucleophile which reacts with the M^{pro} irreversible inhibitors in the S1' pocket (4).

In order to predict the binding mode of SN02 and SN18 with higher accuracy, we ran MD simulations for the ligand-protein complexes resulting from docking for a further 300 ns, where RMSD measurements were monitored for both the ligand molecule and the protein backbone atoms. That was followed by clustering the produced trajectories to select the most frequent protein-ligand conformations in order to analyze the binding modes of the best two compounds.

As shown in Fig. 3a, SN02 seems to be quite stable in the M^{pro} pocket as its RMSD values slightly fluctuating between 1.0–1.5 Å, with the protein backbone also showing a fairly stable tertiary structure (1.5–2.5 Å) throughout the 300-ns MD simulations. In terms of binding, SN02 showed good fitting inside the M^{pro} catalytic pocket (Fig. 3b), spanning over the sub-sites (*i.e.,* S1', S1 and S2). Furthermore, SN02 was able to make several hydrogen bond interactions through its hydroxyl groups with three different residues (*i.e.,* Cys44 and Phe140) located in the S1 and S2 pockets (Fig. 3c). However, the other two polar groups have been placed deep inside the pocket with no clear role in the protein-ligand binding, which seems to be responsible for its high EGB value (55.2 kcal mol^{-1}, Table II). Hence, elimination or replacement of these unnecessary polar groups should be considered in the future to reduce the SN02 desolvation penalty and improve its binding affinity towards the M^{pro} pocket.
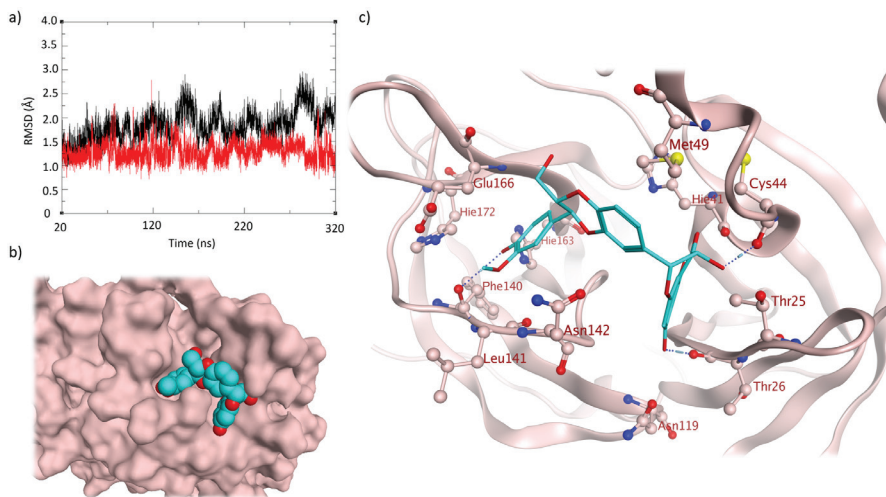
Fig. 3. a) The RMSD values of SN02 (red) and the protein backbone (black) over the 300 ns MD simulation, b) the shape fitting of the ligand molecule inside the M^{pro} active site along with c) the 3D binding mode of the ligand-protein complex.
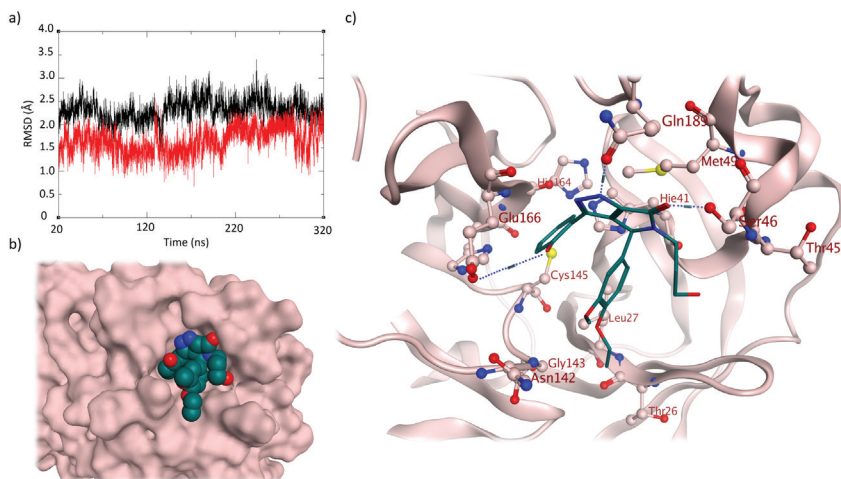

Fig. 4. a) The RMSD values of SN18 (red) and the protein backbone (black) over the 300 ns MD simulation, b) the shape fitting of the ligand molecule inside the M^{pro} active site along with c) the 3D binding mode of the ligand-protein complex.

On the other hand, SN18 seems to be less complicated as it possesses a fewer number of chiral centers compared to the SN02 structure, which introduces this compound as a more attractive candidate for drug-like M^{pro} inhibition. As shown in Fig. 4a, SN18 showed to be stable inside the M^{pro} binding site, with an average RMSD value of 1.7 Å. The com-

pound shows nice complementarity with the target enzyme in terms of both shape and interactions, spanning over four subsites S1′, S1, S2 and S4 (Figs. 4b and 4c). SN18 was able to use its dihydropyrrolopyrazol ring system as an anchor in order to insert two aromatic side-chains along with the aliphatic group into the appropriate sub-pockets in the M^{pro} cavity (Fig. 4c). The anchor in this case is stabilized by forming two strong hydrogen bonds with the Ser46 and Gln189 residues in the S2 and S4 pockets, resp. One more hydrogen bond is formed between the SN19 phenol group and Glu166 side-chain in the S1 pocket. The other polar group, alcohol, present in this structure is partially solvated which probably contributed to the less penalty paid by SN18 (EGB = 50.9 kcal mol^{-1}) compared to SN02 (55.2 kcal mol^{-1}, Table II). Furthermore, these polar groups can be utilized in the future as polar handles for additional binding groups and hence will help in the lead optimization process. To sum up, should SN18 proven to be experimentally active, it can act as a starting point for future anti-COVID-19 drugs especially since it has a simple structure with lead-like characteristics, providing big room for structural modification and activity improvement in the future.

CONCLUSIONS

SARS-CoV-2 main protease protein (M^{pro}) was evaluated as a COVID-19 drug target in this project. The protein was subject to a virtual screening study by the commercially available database (TimTec). The resultant top 21 compounds that had a potential robust binding mode with the viral protein were selected from various chemical clusters. To further enhance the docking results, molecular dynamic studies and MM-GBSA calculations were performed to reassure the docking result of the ligands and predict their binding energies with the target enzyme. Accordingly, four potential inhibitors were identified (SN02, SN18, SN19, SN20) that showed interesting binding modes, along with the lowest binding energies of all ligands. Finally, MD simulations for the ligand-protein complexes of SN02 and SN18 (trihydroxychroman and pyrazolone derivatives) were run for further 300 ns and showed high stability inside the M^{pro} pocket. These results could be the starting point for structure-activity relationship studies for the trihydroxychroman and pyrazolone derivatives as potential inhibitors of SARS-CoV-2 M^{pro} activity.

Abbreviations, acronyms, symbols. – EGB – polar contribution to the solvation free energy, EEL – electrostatic energy, ESURF – nonpolar contribution to the solvation free energy, MD – molecular dynamics, MM-GBSA – molecular mechanics-generalize born surface area, M^{pro} – main protease, SD – standard deviation, VDWAALS – van der Waals contribution, VS – virtual screening.

REFERENCES

1. R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Y. Bi, X. Ma, F. Zhan, L Wang, T. Hu, H. Zhou, Z Hu, W. Zhou, L. Zhao, J. Chen, Y. Meng, J.Wang, Y. Lin, J. Yuan, Z. Xie, J. Ma, W. J. Liu, D. Wang, W. Xu, E. C. Holmes, G. F. Gao, G. Wu, W. Chen, W. Shi and W. Tan, Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, *Lancet* **395** (2020) 565–574; https://doi.org/10.1016/S0140-6736(20)30251-8

2. M. Yuan, N. C. Wu, X. Zhu, C. D. Lee, R. T. Y. So, H. Lv, C. K. P. Mok and I. A. Wilson, A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV, *Science* **368** (2020) 630–633; https://doi.org/10.1126%2Fscience.abb7269

3. A. Wu, Y. Peng, B. Huang, X. Ding, X. Wang, P. Niu, J. Meng, Z. Zhu, Z. Zhang, J. Wang, J. Sheng, L. Quan, Z. Xia, W. Tan, G. Cheng and T. Jiang, Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China, *Cell Host Microbe* **27** (2020) 325–328; https://doi.org/10.1016/j.chom.2020.02.001

4. Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L. W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao and H. Yang, Structure of M(pro) from SARS-CoV-2 and discovery of its inhibitors, *Nature* **582** (2020) 289–293; https://doi.org/10.1038/s41586-020-2223-y

5. A. Shulla, T. Heald-Sargent, G. Subramanya, J. Zhao, S. Perlman and T. Gallagher, A transmembrane serine protease is linked to the severe acute respiratory syndrome coronavirus receptor and activates virus entry, *J. Virol.* **85** (2011) 873–882; https://doi.org/10.1128/JVI.02062-10

6. S. Lam, A. Lombardi and A. Ouanounou, COVID-19: A review of the proposed pharmacological treatments, *Eur. J. Pharmacol.* **886** (2020) Article ID 173451 (11 pages); https://doi.org/10.1016/j.ejphar.2020.173451

7. Chemical Computing Group, *Manual Version 2018.09, Molecular Operating Environment (MOE),* Montreal 2018; http://www.chemcomp.com

8. G. M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju and W. Sherman, Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments, *J. Comp. Aided Mol. Des.* **27** (2013) 221–234; https://doi.org/10.1007/s10822-013-9644-8

9. R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, P. S. Shenkin, Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, *J. Med. Chem.* **47** (2004) 1739–1749; https://doi.org/10.1021/jm0306430

10. J. Wang, W. Wang, P. A. Kollman and D. A. Case, Automatic atom type and bond type perception in molecular mechanical calculations, *J. Mol. Graph. Mod.* **25** (2006) 247–260; https://doi.org/10.1016/j.jmgm.2005.12.005

11. A. Jakalian, D. B. Jack and C. I. Bayly, Fast, efficient generation of high-quality atomic charges, AM1-BCC model: II. Parameterization and validation, *J. Comp. Chem.* **23** (2002) 1623–1641; https://doi.org/10.1002/jcc.10128

12. T. S. Lee, B. K. Allen, T. J. Giese, Z. Guo, P. Li, C. Lin, T. D. McGee, D. A. Pearlman, B. K. Radak, Y. Tao, H. Tsai, H. Xu, W. Sherman and D. M. York, Alchemical binding free energy calculations in AMBER20: Advances and best practices for drug discovery, *J. Chem. Inf. Model.* 2020, in press; https://doi.org/10.1021/acs.jcim.0c00613

13. C. R. Guimarães and M. Cardozo, MM-GB/SA rescoring of docking poses in structure-based lead optimization, *J. Chem. Inf. Model.* **48** (2008) 958–970; https://doi.org/10.1021/ci800004w

14. J. Shao, S. W. Tanner, N. Thompson and T. E. Cheatham, Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms, *J. Chem. Theory Comp.* **3** (2007) 2312–2334; https://doi.org/10.1021/ct700119m

15. W. Dai, B. Zhang, X. M. Jiang, H. Su, J. Li, Y. Zhao, X. Xie, Z. Jin, J. Peng, F. Liu, C. Li, Y. Li, F. Bai, H. Wang, X. Cheng, X. Cen, S. Hu, X. Yang, J. Wang, X. Liu, G. Xiao, H. Jiang, Z. Rao, L. K. Zhang, Y. Xu, H. Yang and H. Liu, Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease, *Science* **368** (2020) 1331–1335; https://doi.org/10.1126/science.abb4489